energies

Article

A Comparative Study on Fuel Consumption Prediction Methods of Heavy-Duty Diesel Trucks Considering 21 Influencing Factors

Jian Gong, Junzhu Shang, Lei Li, Changjian Zhang, Jie He * and Jinhang Ma

School of Transportation, Southeast University, Nanjing 210018, China; timely@topchains.cn (J.G.); 220203401@seu.edu.cn (J.S.); 213162408@seu.edu.cn (L.L.); 230189680@seu.edu.cn (C.Z.); 220194508@seu.edu.cn (J.M.)
* Correspondence: hejie@seu.edu.cn

Abstract: With increasingly prominent environmental problems, controlling automobile exhaust has become essential to the environment. The fuel consumption of transportation is the critical factor that determines exhaust gas. By analyzing the naturalistic driving data of heavy-duty diesel trucks (HDDTs), this paper explored the influence of engine technical state, road features, weather, and temperature conditions on fuel consumption during driving. The detailed process is as follows: Firstly, we collected 1153 naturalistic driving data from 34 HDDTs and made a specific analysis and summary description of the data; secondly, by establishing a binary Logistic regression model, we quantitatively explored the influence of significant factors on the fuel consumption; meanwhile, based on quantitative analysis of factor's effectiveness, this research used several machine learning algorithms (back-propagation neural network, decision tree, and random forest) to build fuel consumption predictors, and compared the prediction performance of different algorithms. The results showed that the prediction accuracy of the decision tree, back-propagation (BP) neural network, and random forest is 81.38%, 83.98%, and 86.58%, respectively. The random forest showed the best performance in predicting. The conclusions can assist transportation companies in formulating driving training strategies and contribute to reducing energy consumption and emissions.

Keywords: environmental protection; fleet management system; heavy-duty diesel trucks; prediction of fuel consumption; binary Logistic regression; machine learning

check for updates

Citation: Gong, J.; Shang, J.; Li, L.; Zhang, C.; He, J.; Ma, J. A Comparative Study on Fuel Consumption Prediction Methods of Heavy-Duty Diesel Trucks Considering 21 Influencing Factors. *Energies* **2021**, *14*, 8106. https://doi.org/10.3390/en14238106

Academic Editor: Ravinesh C Deo

Received: 30 October 2021
Accepted: 30 November 2021
Published: 3 December 2021

1. Introduction

1.1. Background

At present, emphasizing the sustainable development of the environment and energy has become a prerequisite for the operation of industries around the world, especially for the transportation industry. The fuel used by motor vehicles mainly comes from non-renewable energy, such as oil and will produce greenhouse gases and harmful emissions during operation. Therefore, in the context of oil scarcity and atmospheric degradation, sustainable development projects have made the reduction of vehicle fuel consumption a priority. As the main body of road transport, HDDTs generally consume much more fuel than other types of vehicles due to their characteristics, such as large load and long-haul distance. So, it is crucial to analyze the mechanism of the frequent occurrence of high fuel consumption in HDDTs.

The fuel consumption of HDDTs is influenced by various factors [1]. Evaluating the fuel consumption based on mileage is a traditional measure for logistics and transportation companies. However, this method only considers a single factor and has a simple evaluation process, and it cannot reflect the comprehensive impact of human-vehicle-road-environment and other factors. An incomplete understanding of the influencing factors under actual driving conditions will not be conducive to forming a sustainable low-carbon driving system for the entire society.

1.2. Literature Review

1.2.1. Research on Factors Influencing Fuel Consumption

Various factors influence the fuel consumption level of HDDTs in a complex environment. Related studies generally divided the elements into six categories: travel-related, weather-related, vehicle-related, roadway-related, traffic-related, and driver-related factors [2]. These factors were coupled with each other, and they affected the fuel consumption of HDDTs to different degrees.

Hlasny et al. [3] discussed the effects of driver behavior, road conditions, weather, and traffic on fuel consumption of heavy-duty vehicles. They found that changing driving behavior according to these factors before and during driving can reduce fuel consumption. After consulting the relevant literature, it is found that few studies are describing the influencing factors of the fuel consumption of trucks. Therefore, this section also discussed the relevant research of other models. Li et al. [4] analyzed 10-month long-term data collected by private cars in Toyota, and the study explored the relationship between fuel consumption efficiency and drivers' characteristics. They found that some factors have significant influences on the fuel consumption of cars while some elements are almost negligible. Chen et al. [5] developed a mesoscopic fuel consumption estimation model which included predictors that were hardly regarded before, such as the number of lanes and free-flow speed. The results of the study showed that these factors also have an impact on motor vehicle fuel consumption.

In this paper, the factors affecting fuel consumption were divided into four categories: vehicle-related factors, environment-related factors, driving-related factors, and road-related factors. Vehicle-related factors include three types: engine technical state, driving system technical state, and transmission system technical state [6]. The environment-related factors can be divided into five main aspects: average altitude, temperature, humidity, wind, and weather conditions. The differences in drivers' performance on fuel consumption are reflected in both the long-term and the short-term. The long-term performance will be divided explicitly into long-term driving styles, long-term driving habits, and going qualifications. In contrast, short-term performance includes driving styles influenced by weather and date [7]. Road-related factors contain road features and road geometry, such as road curvature and slope, as well as mileage [8]. The categories of influencing factors are clearly shown in Table 1.

Table 1. The categories of influencing factors.

	Factors	Source
Vehicle-related	Engine technical state Driving system technical state Transmission system technical state	[6]
Environment-related	Average altitude Temperature Humidity Wind Weather conditions	[7]
Driving-related	Long-term driving styles Long-term driving habits Going qualifications Driving styles influenced by weather and date	[7]
Road-related	Road features Road geometry	[8]

1.2.2. Research on Fuel Consumption Model

In recent years, scholars have focused on constructing the fuel consumption model under the influence of multiple factors, and the modeling method is gradually diversified with more accuracy.

The commonly used physical models for fuel consumption calculation include vehicle specific power (VSP) model, Virginia Tech microscopic (VT-Micro) model, and comprehensive modal emissions model (CMEM), etc. However, these models have more coefficients and the model calibration process is complicated. Wang et al. [9] developed a VSP-based fuel consumption calculation model using a portable exhaust emission measurement system, but the accuracy range of this model is not high. Xiang et al. [10] established a database containing the actual driving information of heavy trucks, developed an effective model that can quantitatively calculate the driving fuel consumption of vehicles, but this model did not verify the calculation accuracy. J. Wang & Rakha [11] constructed a fuel prediction model of heavy-duty trucks based on Virginia Tech Comprehensive Power-based Fuel consumption Model (VT-CPFM) and calibrated the model with field test data. The results showed that the prediction accuracy of this model was better than that of CMEM.

Machine learning methods have been widely used in fuel consumption analysis with data mining and information technology development. The advantage of exploring fuel consumption prediction accuracy with machine learning models over physical models is that large-scale samples can be processed quickly and efficiently. Although the prediction accuracy may not be as good as physical models, the machine learning algorithms are updated and iterated quickly and have fewer bottlenecks. Ma et al. [12] used the C4.5 decision tree to evaluate the influence of driving behavior on the fuel of urban buses during acceleration. This model had more than 85% prediction accuracy with fewer training samples and strong generalization ability. Du et al. [13] used a BP neural network to establish a fuel consumption prediction model, which analyzed fuel consumption level from the time dimension and space dimension, to comprehensively describe the relationship between fuel consumption and influencing factors. The results showed that the BP neural network established in this article has good performance and is suitable for fuel prediction. Wysocki et al. [14] proposed a method based on an artificial neural network, which achieved the goal of high-precision prediction of diesel engine fuel consumption.

1.2.3. Research on Fleet Management System

Nowadays, most logistics and transportation enterprises have begun to build the internal fleet intelligent management system. Compared with manual recording, this advanced digital monitoring and collection method is more accurate and convenient, and the collected data is more diversified. Current research focuses on how to process and use the data collected by the fleet intelligent management system. Carrying on in-depth analysis of these data and mastering the law of fuel consumption during vehicle operation can reduce the fuel consumption of HDDTs to a certain extent.

In freight transportation, there are many fleet management systems based on remote information monitoring, like Bluetree, Verizon Connect, GPS Insight, and so on [15]. Most transportation enterprises track and collect data through on-board sensors installed on the vehicle. Various data collected by these sensors, such as speed, fuel consumption, real-time positioning, engine temperature, etc., are essentially the transmission of vehicle driving environment, driving status, driver behavior, and other information. Then the collected data is sent back to the enterprise's internal database; that is, the real-time data tracking is completed, and the collected data can be used for subsequent analysis.

On-board diagnostics (OBD) and global positioning systems (GPS) are usually the most frequently used real-time detection equipment in the fleet management system. Walnum and Simonsen [16] made a systematic analysis of the data collected by the fleet management system. They confirmed that different driving modes have different impacts on fuel consumption under other environmental conditions. According to the on-board data recorder, Toledo and Shiftan [17] evaluated the effects of different environments on the driving process to improve driving safety and reduce fuel consumption. Based on the real-time driving and fuel consumption data collected by on-board sensors, Sun et al. [18] estimated the fuel consumption of diesel buses by establishing the corresponding fuel consumption model. Based on the on-board measurement system installed on the vehicle, Guo et al. [19] recorded the fuel consumption and emissions per second. They established the vehicle fuel consumption and emissions model by collecting different types of data many times. Similarly, Faria et al. [20] used a commercial vehicle data recorder to collect a large number of actual vehicle operation data to more comprehensively study driving behavior's impact on fuel consumption.

1.3. Research Objectives and Innovation

Through the above research, we can see that fleet management systems have become extremely frequent with the popularity of big data. Many studies use on-board sensors to record real-time vehicle operation data. However, large transportation enterprises still have blind areas of knowledge on how to use the data collected by the fleet management system effectively. The collected data has not been fully and concretely used and can't play to the maximum value. The capital and efforts invested by the enterprises are not equal to the returns. In addition, the existing fuel consumption prediction approaches also have some limitations. The fuel consumption models used in the above studies are independent; they can't be reflected in the same dimensional scenario, so it is difficult to decide which approach is more suitable.

This research took a small and medium-sized logistics enterprise engaged in national and provincial trunk road logistics transportation as the research case to solve these problems. By making full use of the actual data collected by the fleet management system, this paper constructed the binary Logistic regression model, the neural network model, the decision tree model, and the random forest model, respectively, to analyze the fuel consumption of HDDTs. The conclusion of this paper is helpful for enterprises to formulate fuel-saving strategies from the perspective of drivers' behavior training and has far-reaching significance for reducing exhaust gas and environmental pollution.

The main innovations of this paper can be summarized as follows: First, it shows how to make full use of the real-time operation data collected by the fleet management system inside the enterprise to analyze the fuel consumption of vehicles. Second, this paper discusses the influence of different factors of humans, vehicles, roads, and the environment on the fuel consumption of HDDTs. Third, the performance of different fuel consumption prediction methods in HDDTs' fuel consumption analysis is compared.

A brief overview of this paper is now provided. The experimental models used in this research will be briefly described in the next section, followed by an overview of the data. Then through the binary Logistic regression model, the influencing factors related to the fuel consumption of HDDTs are analyzed. The prediction model suitable for the specific situation of this paper is selected through the comparison of prediction accuracy among machine learning models. The simple structural framework of this paper is shown in Figure 1.

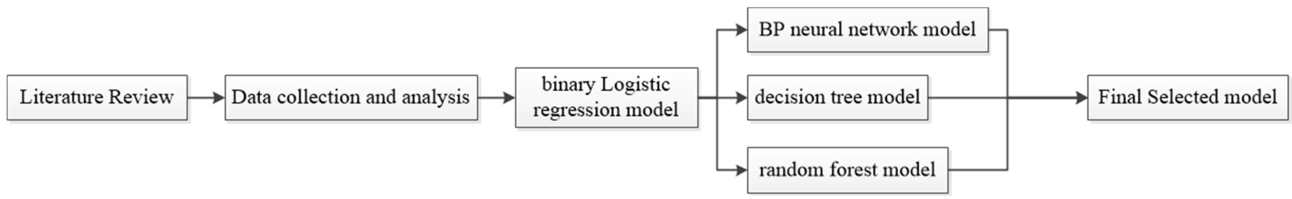

Figure 1. The simple structural framework.

2. Data and Method

2.1. Data

2.1.1. Data Source

This paper took TopChains International Logistics Co., Ltd. (Shenzhen, China) as a research case to study the fuel consumption characteristics of HDDTs. TopChains International Logistics Co., Ltd. is a modern small and medium-sized logistics company. The HDDTs studied in this paper are the Shandeka SITRAK heavy semi-trailer tractors belonging to Sinotruk corporation (Jinan, China). The specific parameters are shown in Table 2.

Table 2. Parameters of Shandeka SITRAK HDDTs.

Parameter Type	Parameter Value	Parameter Type	Parameter Value
Drive form	4X2 or 6X2R	Vehicle weight	8.54 tons
Engine	Sinotruk MC13.54-50	Total mass	25 tons
Maximum horsepower	540 horsepower	Fuel type	diesel
Emission standards	National five	Number of passengers	3 people
Gearbox	ZF16S2530 TO	Displacement	12.419 L

The data were collected from the SINOTRUK Intelligent Platform of TopChains International Logistics Co., Ltd., which is applied to HDDTs' management and monitoring. This platform has entered the information of the company's 34 HDDTs engaged in the general transportation and logistics business. It performs real-time positioning monitoring and data back statistics through the vehicle network terminal management system. The concrete operation interface of the platform is shown in Figure 2.

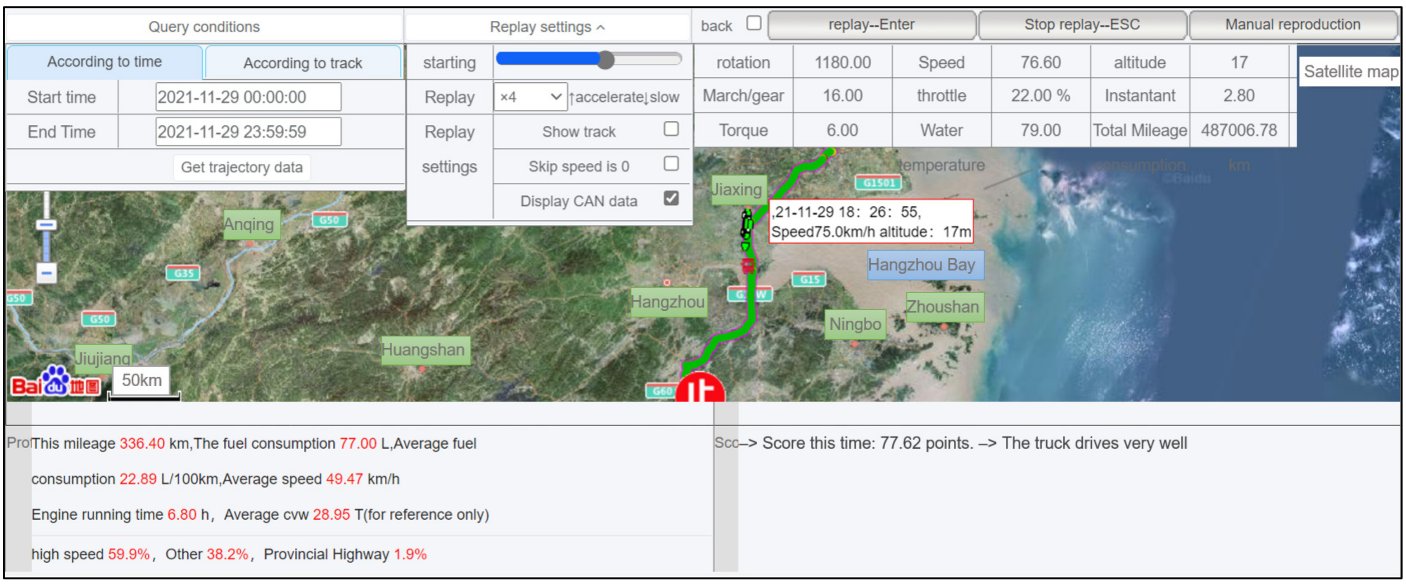

Figure 2. Operation page of SINOTRUK Intelligent Platform.

The platform obtains data, such as the positioning information and the technical status of vehicles through the GPRS system, which comprises vehicle hardware of the vehicle terminal, satellite positioning system, and mobile network. It transmits the data back to the mobile gateway and enters them into the intelligent library of SINOTRUK for reflection on the platform. The specific principle is shown in Figure 3.

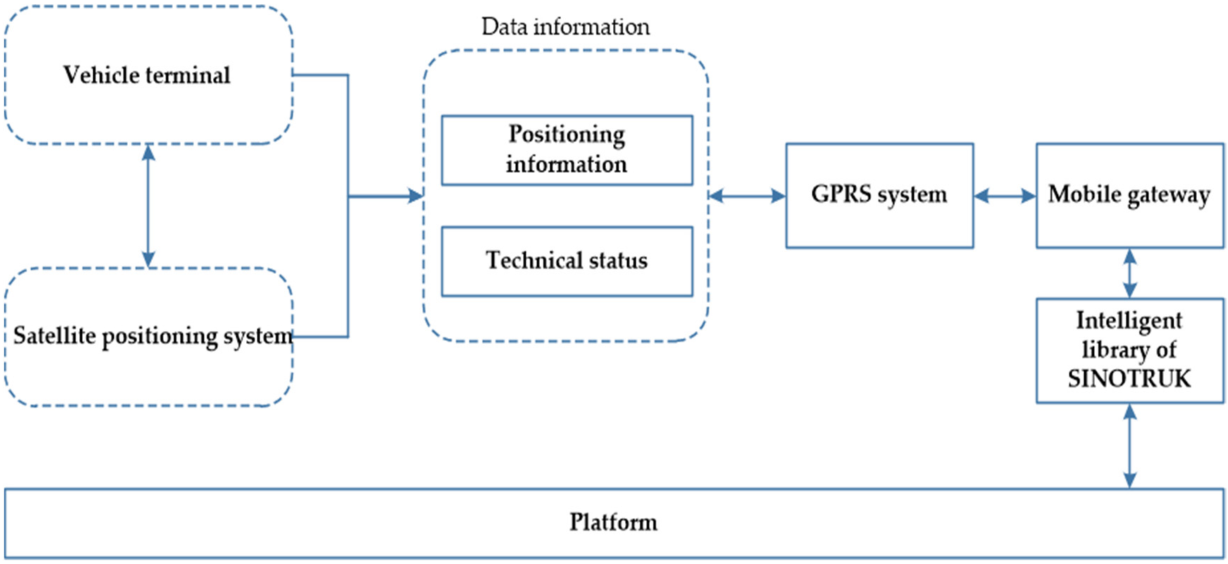

Figure 3. Schematic diagram of SINOTRUK Intelligent Platform.

2.1.2. Data Processing

(1) Specific data types

Through mining and analyzing the collected data, it was found that the vehicle-related factors affecting the fuel consumption of HDDTs include seven continuous variables: weight, average rotating velocity, standard deviation rotating velocity, average velocity, standard deviation velocity, economic rotating velocity ratio, and non-economic rotating velocity ratio. These were obtained directly through the single truck operating status information of the Sinotruk Intelligent Platform.

There were five variables in environment-related factors. The average altitude and altitude change were continuous variables, and another three discrete variables included temperature, weather, and holiday. The weather data were represented by precipitation intensity. A truck traveled through several cities in one day, and the temperature and climate of each city were different. According to the replay information of the vehicle's trajectory, the city where the truck was located each time could be known. Each area's temperature and precipitation intensity could be obtained by checking the weather data obtained from the China Meteorological Administration. Knowing the length of stay and the corresponding temperature and precipitation in each stop city, the weighted temperature, and rain of the day can be obtained by Equations (1) and (2).

$$T = \frac{\sum(T_n \times M_n)}{\sum M_n} \tag{1}$$

$$P = \frac{\sum(P_n \times M_n)}{\sum M_n} \tag{2}$$

In these formulas, T represents temperature, and P represents precipitation intensity, while M_n represents the time of a truck driving in the nth city, and T_n, P_n are the temperature and precipitation intensity when the vehicle is driving in that city.

The driving-related factors mainly included neutral taxiing ratio, gear taxiing ratio, idle speed ratio, and parking time ratio; the road-related factors included the mileage and the road classifications of driving, such as freeway, national highway, province highway, and other ordinary roads. All these were continuous variables.

(2) Data standardization

To reasonably determine the fuel consumption threshold (high, normal), this paper used the quartile of fuel consumption per 100 km to calibrate it discretely. When the fuel consumption per 100 km was higher than the overall upper quartile (that means the fuel consumption per 100 km is higher than 75% of the fuel consumption level), it would be labeled as the high fuel consumption. The rest would be judged as normal fuel consumption.

The study included four discrete variables: weather, temperature, holiday, and fuel consumption per 100 km. The statistics of discrete variables are primarily recorded in words, and data mining processing software can't identify and analyze these records. It is necessary to standardize the discrete data and convert them into coding symbols that the data mining processing software can recognize. The standardized processing results are shown in Table 3.

Table 3. Standardized treatment for discrete variables.

Discrete Data Name	Classification Description	Standardized Value
Holiday	Yes	1
	No	0
Temperature	Under 10 °C	0
	11–15 °C	1
	16–20 °C	2
	21–25 °C	3
	25–30 °C	4
	More than 30 °C	5
Weather	No rain	0
	Precipitation 1–8 mm	1
	Precipitation 10–20 mm	2
Fuel consumption per 100 km	Normal fuel consumption	0
	High fuel consumption	1

(3) Data summary statistics

In a word, the data set contained a total of 1153 pieces of data. The data set included 21 independent variables, and the dependent variable was fuel consumption per 100 km. The summary description results are shown in Table 4.

Table 4. The summary of HDDTs' naturalistic driving data.

Variable Name (Type)	Definition
Driving Characteristics	
Neutral taxiing ratio(continuous)	Percentage of truck driving time with no engine load during a trip
Gear taxiing ratio(continuous)	Percentage of truck driving time with engine load during a trip
Idle speed ratio(continuous)	Percentage of time spent idling during a trip
Parking time ratio(continuous)	Percentage of time spent parking during a trip

Table 4. *Cont.*

Variable Name (Type)	Definition
	Environment Characteristics
Average altitude(continuous)	Average altitude per trip/(100 m)
Altitude change(continuous)	The change of altitude per trip/(100 m)
Holiday(discrete)	A discrete variable indicating whether the driving day is a holiday
Temperature(discrete)	A discrete variable indicating outdoor temperature while driving
Weather(discrete)	A discrete variable indicating weather while driving, expressed in precipitation in this paper
	Vehicle Characteristics
Weight(continuous)	Average cargo weight per trip/(ton)
Average rotating velocity(continuous)	Average engine rotating velocity per trip/(100 r/min)
Standard deviation rotating velocity(continuous)	The standard deviation of engine rotating velocity per trip
Average velocity(continuous)	Average speed per trip/(km/h)
Standard deviation velocity(continuous)	The standard deviation of speed per trip
Economic rotating velocity ratio(continuous)	Percentage of truck driving time in the economic rotating velocity range during a trip
Non-economic rotating velocity ratio(continuous)	Percentage of truck driving time in the non-economic rotating velocity range during a trip
	Road Characteristics
Freeway(continuous)	Percentage of distance the truck travels on freeways during a trip
National road(continuous)	Percentage of distance the truck travels on ordinary national roads during a trip
Provincial road(continuous)	Percentage of distance the truck travels on ordinary provincial roads during a trip
Other ordinary roads(continuous)	Percentage of distance the truck travels on other low-grade roads during a trip
Mileage(continuous)	Mileage during a trip
Fuel consumption(discrete)	Fuel consumption per hundred kilometers for each trip

2.2. Methodology

The methods used in this study to handle the above data are as follows:

2.2.1. Binary Logistic Regression

Logistic regression is a method of modeling and analysis based on the fitting of nonlinear relationships, which is used to study the interactions and dependencies between independent and dependent variables. The independent variables can be either continuous or discrete or a combination of the two. It does not require that the residuals of the variables satisfy normal distribution, nor does it need a linear correlation between the variables, so its results are more objective [21]. Through Logistic analysis, the influencing weights of dependent variables under the coupling of independent variables can be obtained. The significant influencing factors can be extracted, and the probability of dependent variables can be predicted based on the coefficients.

Logistic regression can be divided into two models: binary classification and multinomial classification. In the binary classification model, the dependent variable can only take 0 or 1, while in the multinomial classifications model, the dependent variable can be divided into several categories. The binary Logistic model would explore the relationship between various factors and the fuel used in this study. The basic form of binary Logistic model is as follows:

$$\mathrm{Ln}\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n, n = 1, 2 \ldots \quad (3)$$

Among them, x_1, $x_2 \ldots\ x_n$ are explanatory variables, while β_1, $\beta_2 \ldots\ \beta_n$ represent the regression coefficients. P stands for the probability that the classification result is 1 [22]. In this paper, the dependent variables are divided into two categories (high fuel consumption and normal fuel consumption). 1 represents high fuel consumption, and 0 represents normal fuel consumption. And x_1, $x_2 \ldots\ x_n$, in this paper, mean the factors which will affect the fuel consumption of HDDTs.

2.2.2. BP Neural Network

BP artificial neural network is a multilayer feedforward neural network based on error back-propagation. It has the characteristics of nonlinear mapping, adaptive learning, and robust fault tolerance and is suitable for dealing with nonlinear problems. It is mainly composed of an input layer, one or more hidden layers, and an output layer, and each layer is formed of several neurons. The output values of each node are determined by input values, excitation function, and threshold values [23].

The algorithm includes two processes: forward propagation of signal and back-propagation of error [24]. In forwarding propagation, the input values act on the output nodes through hidden layers and generate the output signals through nonlinear transformation. Then through the process of error back-propagation, the connection weights are continuously modified until the error signals are minimized and the output values meet the required accuracy. The most basic three-layer BP neural network structure is shown in Figure 4.

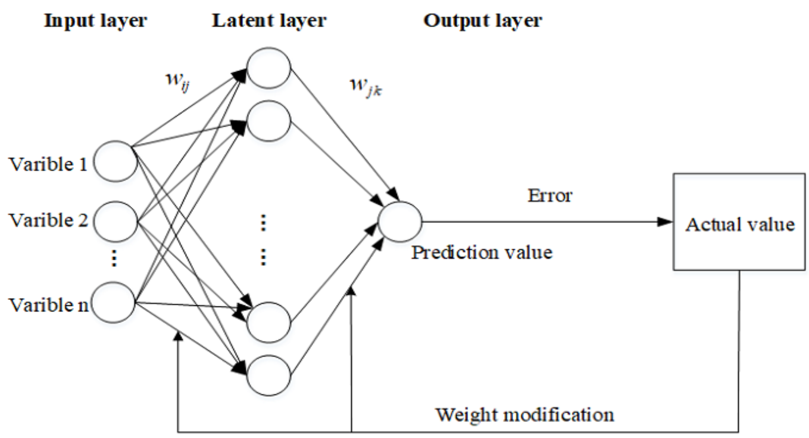

Figure 4. Three-layer BP neural network structure diagram [25].

In Figure 4, W_{ij} represents the connection weight between the input and the hidden layer, W_{jk} describes the connection weight between the hidden and the output layer. In addition, for each neuron in the output layer, there is also a threshold that regulates the level of neuron excitement. The weight and threshold are initially given arbitrary values. Then after a cyclic process of back-propagation, they are continuously updated according to the error signals until they meet the accuracy requirements [25].

In this research, input values are independent variables that have vital impacts on fuel consumption of HDDTs; output values are predicted fuel consumption. The output values will be compared with the actual values to judge whether the performance of the BP neural network model is suitable for the fuel consumption prediction of HDDTs under a natural driving state in this study.

2.2.3. Decision Tree

The decision tree is a commonly used classification machine learning method. It can be divided into the classification and regression trees: the former outputs classifications, and the latter outputs numerical values.

CART decision tree belongs to the classification regression tree; its binary tree structure is intuitive and stable. It is suitable for processing large quantities of data, and the speed is fast. The constructed decision tree starts from the root-node and divides the complex data set into specific types by generating gradually refined branches. The specific types are represented by leaf-nodes [26]. As such, the decision tree can categorize the data intuitively. However, during the modeling process of the CART classification tree, if the growth of branches is not restricted, the decision tree will be too complex, and the accuracy of the model will decrease. To improve the precision of the decision tree and prevent the overfitting of the model, the decision tree should be pruned.

The main processes of the CART decision tree algorithm are branching and pruning. According to the principle of minimum Gini coefficient, a feature is selected from the points of the given training sample set as the splitting standard, then generate the corresponding child-nodes recursively from top to bottom according to it, until the data set can no longer be divided. The pruning process of the decision tree can be divided into pre-pruning and post-pruning. This paper chose the method of post-pruning: if the tree grows too large, the pruning criterion is used to reduce the node to a more suitable size according to the error rate. In this way, the problem of overfitting the decision tree is avoided [27].

In this paper, the influencing factors related to fuel consumption of HDDTs are the selected segmentation standards. Then it is divided layer by layer from the root-node according to the standards. The final output leaf-nodes are the classification result values: high fuel consumption and normal fuel consumption.

2.2.4. Random Forest

Random forest is a machine learning algorithm published by Breiman [28]. It contains multiple decision trees trained by bagging integration technology, and the final learning result is voted by the output results of all decision trees. Its basic unit is the decision tree, and its essence belongs to the integrated learning method. The random forest algorithm has better prediction accuracy for high-dimensional problems and generally does not suffer from overfitting [29].

The random forest algorithm has two main links: the growth of the decision tree and the voting process. When constructing the classification tree, the random forest will take the bootstrap sampling method to randomly select the sample set from the sample data and repeat k times to form a new training sample set. Then randomly choose m ($m < M$) features from M input features and choose one element from m features for the branch growth according to the principle of minimum node impurity. Repeating the above process to generate all branches until all the attributes have been used or the tree can accurately classify the training set. It should be noted that m remains constant throughout the process. To minimize the impurity of each node, the regular pruning operation is not performed. In general, hundreds to thousands of classification trees are generated randomly in a random forest algorithm, and the final output value will be determined by voting [30].

Similar to the decision tree algorithm, the random forest model in this paper also recurses the tree with the influencing factors of fuel consumption as eigenvalues, and finally generates the classification result values of fuel consumption of HDDTs.

In the above methods, the binary the Logistic model is used to qualitatively explain the relationship between the influencing factors and the fuel consumption of HDDTs, and by comparing the actual fuel consumption value and the predicted fuel consumption value (that is, the prediction accuracy) of the BP neural network model, CART decision tree model and random forest model, the prediction model most suitable for the driving process of HDDTs in this study is selected.

3. Modeling Results and Discussions

In this section, the binary Logistic regression model focused on explanation was combined with the BP neural network model, the decision tree model, and the random forest model focused on prediction to extract significant influencing factors and discuss the occurrence mechanism of high fuel consumption in a natural driving environment. Then, the result of classification prediction tests of each model was compared.

3.1. Binary Logistic Regression Model

(1) Collinearity diagnosis of variables

It is necessary to diagnose the collinearity of the independent variables and eliminate the variables with collinearity before performing binary Logistic regression to make the model regression results more accurate and reliable. Collinearity diagnosis was made for 21 variables in the influencing factors of the fuel consumption of HDDTs [31]. The diagnostic coefficient results are shown in Table 5.

Table 5. Collinearity diagnosis coefficients.

	VIF	1/VIF		VIF	1/VIF
Weight	2.845	0.352	Idle speed ratio	90,579.430	0.000
Freeway	991.271	0.001	Non-economic rotating velocity ratio	87,492.734	0.000
National road	264.490	0.004	Parking time ratio	602,7043.000	0.000
Provincial road	315.954	0.003	Average altitude	2.781	0.360
Other ordinary roads	535.602	0.002	Altitude change	3.331	0.300
Mileage	39.075	0.026	1.Holiday	1.103	0.906
Average rotating velocity	4.521	0.221	1.Temperature	1.998	0.501
Standard deviation rotating velocity	5.499	0.182	2.Temperature	1.996	0.501
Average velocity	5.751	0.174	3.Temperature	1.707	0.586
Standard deviation velocity	3.366	0.297	4.Temperature	1.425	0.702
Economic rotating velocity ratio	4,832,594.500	0.000	5.Temperature	1.119	0.894
Neutral taxiing ratio	16,775.566	0.000	1.Weather	1.104	0.905
Gear taxiing ratio	7721.713	0.000	2.Weather	1.039	0.962
Mean VIF	425,553.600				

When the variance inflation factor (VIF) > 10, there may be multicollinearity between variables [32]. It can be seen from the table that the data have serious collinearity. The factors with VIF more outstanding than ten are removed here, and the following elements: weight, average rotating velocity, standard deviation rotating velocity, average velocity, standard deviation velocity, average altitude, altitude change, holiday, temperature, and weather will be included in the binary Logistic regression model.

(2) Binary Logistic regression model

The ten factors filtered out above were entered into the STATA software as explanatory variables for binary Logistic regression analysis, and the regression results are shown in Table 6.

Table 6. Binary Logistic regression model results.

Fuel Consumption	Coef.	St.Err.	t-Value	*p*-Value	95% Conf	Interval	Sig
Weight	1.617	0.055	14.020	0.000	1.512	1.730	***
Average rotating velocity	0.989	0.002	−6.320	0.000	0.985	0.992	***
Standard deviation rotating velocity	0.984	0.005	−3.410	0.001	0.975	0.993	***
Average velocity	0.769	0.017	−11.780	0.000	0.736	0.803	***
Standard deviation velocity	0.887	0.034	−3.130	0.002	0.823	0.956	***
Average altitude	1.002	0.001	2.700	0.007	1.001	1.004	***
Altitude change	0.999	0.000	−1.330	0.184	0.998	1.000	
0.Holiday	*base*						
1.Holiday	0.923	0.586	−0.130	0.900	0.266	3.206	
0.Temperature	*base*						
1.Temperature	0.796	0.212	−0.860	0.392	0.473	1.341	
2.Temperature	1.000	0.291	−0.000	1.000	0.566	1.768	
3.Temperature	1.777	0.603	1.700	0.090	0.914	3.455	*
4.Temperature	1.240	0.605	0.440	0.659	0.477	3.227	
5.Temperature	18.100	14.03	3.740	0.000	3.962	82.698	***
0.Weather	*base*						
1.Weather	0.552	0.154	−2.130	0.033	0.320	0.954	**
2.Weather	0.935	0.931	−0.070	0.946	0.133	6.587	
Constant	1,184,552.6	2,242,029.9	7.39	0	29,004.515	48,377,460	***

Comment: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

It can be seen from the above table that the eight variables of weight, average rotating velocity, standard deviation rotating velocity, average velocity, standard deviation velocity, average altitude, temperature, and weather are significant under the 95% confidence level. What's more, the change of the probability value of the explained variable can be judged from the coefficient values of the explanatory variables. Still, the specific weight of the change can't be seen intuitively. Therefore, the margins command was utilized to analyze its marginal effects, and the results are shown in Table 7.

Table 7. Results of marginal effects.

Variables	Odds Ratio	Std.Err.	z	$p > z$	95% Conf.	Interval
Weight	0.046	0.004	12.640	0.000 ***	0.039	0.054
Average rotating velocity	−0.001	0.000	−6.200	0.000 ***	−0.001	−0.001
Standard deviation rotating velocity	−0.002	0.000	−3.420	0.001 ***	−0.002	−0.001
Average velocity	−0.025	0.002	−11.000	0.000 ***	−0.030	−0.021
Standard deviation velocity	−0.012	0.004	−3.060	0.002 ***	−0.019	−0.004
Average altitude	0.001	0.000	2.700	0.007 ***	0.000	0.000
Altitude change	−0.000	0.000	−1.320	0.187	−0.000	0.000
1.Holiday	−0.007	0.058	−0.130	0.897	−0.121	0.106
1.Temperature	−0.019	0.023	−0.830	0.405	−0.065	0.026
2.Temperature	−0.000	0.027	0.000	1.000	−0.053	0.053
3.Temperature	0.067	0.042	1.600	0.111	−0.015	0.149
4.Temperature	0.022	0.052	0.420	0.675	−0.080	0.123
5.Temperature	0.573	0.163	3.510	0.000 ***	0.253	0.892
1.Weather	−0.049	0.020	−2.480	0.013 **	−0.088	−0.010
2.Weather	−0.007	0.098	−0.070	0.944	−0.200	0.186

Comment: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The influential variables introduced in this binary Logistic regression model are weight, average rotating velocity, standard deviation rotating velocity, average velocity, standard deviation velocity, average altitude, altitude change, holiday, temperature, and weather. The results of the regression analysis are shown in Tables 6 and 7. The Odds Ratio (OR) value and p value are provided as model performance indicators, and the Odds Ratio value represents the ratio of the probability of occurrence to the probability of non-occurrence in the study group divided by the ratio in the control group [33]. The description of the binary Logistic regression results based in Table 7 is as follows.

For weight, the OR value is 0.046, indicating that when the average value of weight increases by 1 unit, the probability that the fuel consumption of HDDTs exhibits a high fuel consumption level increases by 4.6%. This is easy to comprehend. When the vehicle is heavily loaded, the driving process often requires more fuel [34]. Drivers must not overload when driving HDDTs to transport goods. It will not only lead to increased fuel consumption but also violate safe driving regulations.

Regarding average rotating velocity, the OR value is −0.001, indicating that when the mean of the average rotating velocity increases by one unit, the probability that the fuel consumption of HDDTs exhibits a high fuel consumption level decreases by 0.1%. A similar finding has been found in previous studies. When the gear is fixed, there is a negative correlation between engine speed and fuel consumption [35]. That means for the drivers, it is necessary to avoid the occurrence of low gear and high speed when driving. Generally speaking, drivers who drive HDDTs for cargo transportation are professional and experienced. They usually drive vehicles with matching gears.

The OR value of standard deviation rotating velocity is −0.002, indicating that when the average value of standard deviation rotating velocity increases by one unit, the probability that the fuel consumption of HDDTs exhibits a high fuel consumption level decreases by 0.2%.

For average velocity, the OR value is −0.025, indicating that when the mean of the average vehicle speed increases by 1 unit, the probability that the fuel consumption of HDDTs shows a high fuel consumption level decreases by 2.5%. This signifies that when the average speed increases, the fuel consumption does not increase significantly but shows a downward trend [36]. Of course, this does not mean that the higher the speed of trucks, the lower the fuel consumption, but that when the gears are matched, the drivers maintain a high and stable speed, the corresponding fuel consumption may be steady.

As for standard deviation velocity, the OR value is −0.012, indicating that when the average value of standard deviation velocity increases by 1 unit, the probability that the fuel consumption of HDDTs shows a high fuel consumption level decreases by 1.2%.

Regarding Average altitude, the OR value is 0.001, indicating that when the mean of the average altitude increases by 1 unit, the probability that the fuel consumption of HDDTs exhibits a high fuel consumption level increases by 0.1%. According to relevant studies, vehicles' fuel usage will increase with altitude [37].

For Temperature, the significant influencing factor at this time is that the temperature is more than 30 °C. The OR value is 0.573, indicating that under other conditions unchanged, when the driving temperature of HDDTs is above 30 °C, the average value of temperature increases by 1 unit, the probability of fuel consumption showing a high fuel consumption level increases by 57.3% compared to when the temperature is below 10 °C. Usually, when the temperature is higher than 30 °C, the driver chooses to turn on the air conditioner, significantly increasing fuel consumption. In short, hot weather will make the cars use more fuel [38].

For weather, the significant influencing factor is that the precipitation is 1–8 mm. The OR value is −0.049, indicating that under other conditions unchanged when the precipitation of HDDTs is within 1–8 mm during driving, the average value of weather increases by 1 unit, the probability of fuel consumption showing a high fuel consumption level decreases by 4.9% compared to the case of no rain. We can find that a small amount of precipitation will help the vehicle to save fuel. This is slightly different from previous research's results [39]. The reason for this may be that light rain makes drivers more attentive. The driving process becomes stable, offsetting the positive impact between precipitation and fuel consumption.

The analysis of the above results indicates eight factors that significantly affect the fuel consumption of HDDTs in this study. It can be seen they are factors related to the driving state of vehicles and the environment. For the drivers, the increase in fuel consumption caused by factors related to the driving state of vehicles can be avoided through the drivers' intentional learning and attention, such as prohibiting overload and avoiding the vehicle's driving state in low gear and high speed as far as possible. However, the impact of environmental factors on fuel consumption is difficult to be eliminated by manpower.

From the above modeling process of binary Logistic regression, we could see that this method has higher requirements for data distribution characteristics, leading to a tedious testing process and more elimination factors. The overall model focused on interpretability. Therefore, machine learning methods may be used in the following subsections to build models focused on prediction.

3.2. Machine Learning

3.2.1. Model Training

The author used the BP neural network, CART decision tree, and random forest to establish the prediction models of fuel used. Although the binary Logistics regression model had been established in the previous section and the critical factors had been screened out, it was observed that 21 factors in the data set all affect the generation mechanism of fuel consumption. Therefore, all the factors were still traversed as the inputs of the models. A simple logical framework diagram of input/output variables is shown in Figure 5.

All data were randomly sampled in this paper. The data set was divided into approximately 80%/20% for training and testing to ensure proper models training. Meanwhile, to eliminate the calculation bias caused by the randomness of the result of each model, ten tests were repeated for all models, respectively, and the average of the results of the ten tests was taken as the final test result. The specific process of one test is shown below.

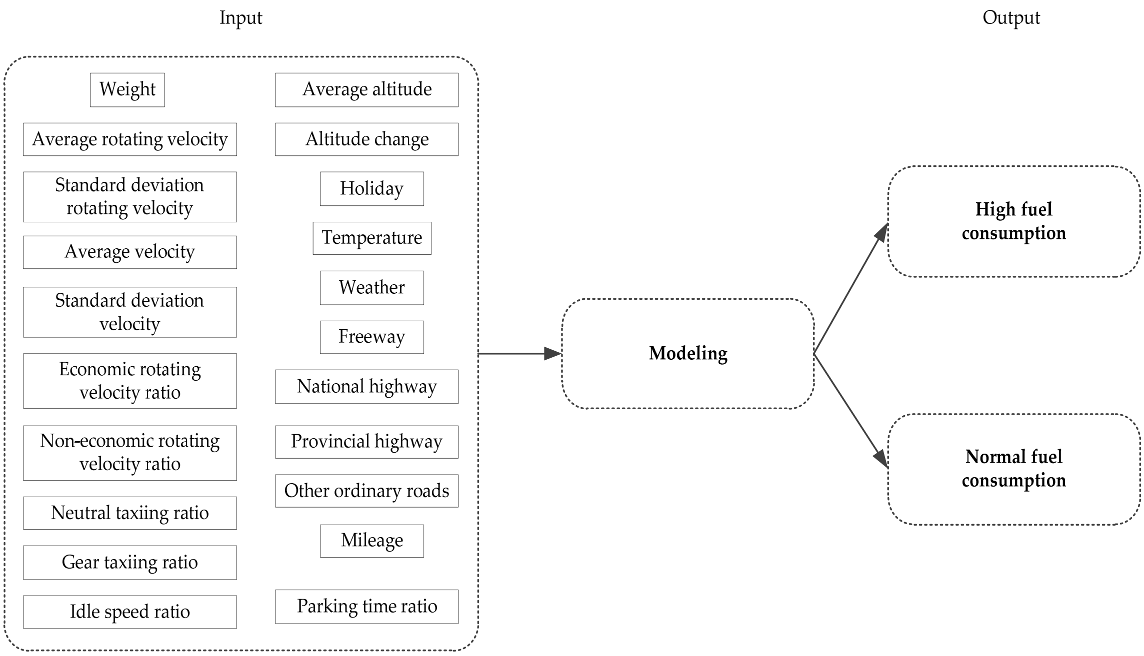

Figure 5. The simple logical framework of input/output variables.

We established the BP neural network, including an input layer, a hidden layer, and an output layer. The package called "nnet" in the R Programming Language for modeling was utilized. The number of input nodes was 21, and the output node was fuel consumption classifications. What needs attention is the number of hidden layer neurons. After multiple iterations of training, the number of hidden layer neurons was determined to be ten according to the minimum prediction error of the training result [40].

For the prediction model of the CART decision tree, pruning is required when building. After establishing a fully-grown decision tree in an overfitting state, this research chose the post-pruning method to prune the tree from bottom to top. The pruning process was usually carried out by setting the threshold of the complexity parameter (cp). The pruned decision tree classifier is obtained according to the cp value corresponding to the minimum relative error [41]. It can be seen from Table 8 that the best choice for cp in the decision tree is 0.018242.

Table 8. Decision tree model performance (cp = 0.01).

Number	cp	nsplit	rel Error	xerror	xstd
1	0.064677	0	1.00000	1.00000	0.062374
2	0.044776	3	0.80597	0.93035	0.060744
3	0.024876	4	0.76119	0.95025	0.061223
4	0.018242	6	0.71144	0.92537	0.060623
5	0.014925	12	0.58209	0.92783	0.060822
6	0.010000	15	0.53731	0.93532	0.060865

xerror: x-val relative error.

In the fuel-use model based on random forest, the main parameters affected the model, including the number of variables randomly sampled when constructing decision tree branches and the number of decision trees. By manually searching for parameters, the number of the variables was 21 in this model. The number of trees can be judged according to the out-of-bag (OOB) error of the model. When the OOB error stabilizes, it is the best choice [42]. As shown in Figure 6, when the value is 200, the OOB error tends to be stabilized [43].

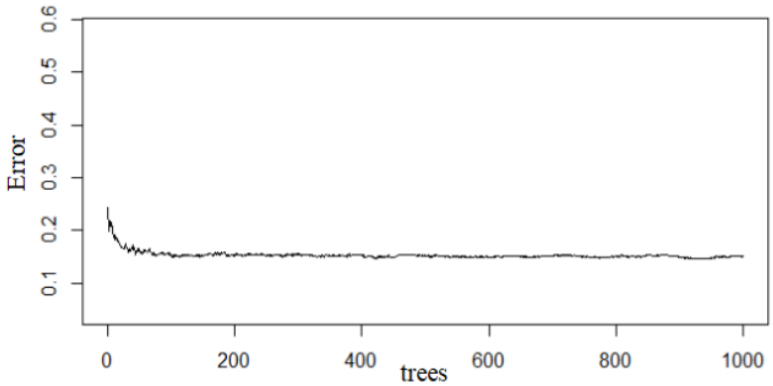

Figure 6. The relationship between the number of decision trees and the error rate.

3.2.2. Model Results and Comparison Analysis

Run the trained models. We selected the best one by comparing the prediction accuracy of each model for fuel consumption. The test result of the above BP neural network model, CART decision tree model, and random forest model was summarized and compared, illustrated in Figure 7.

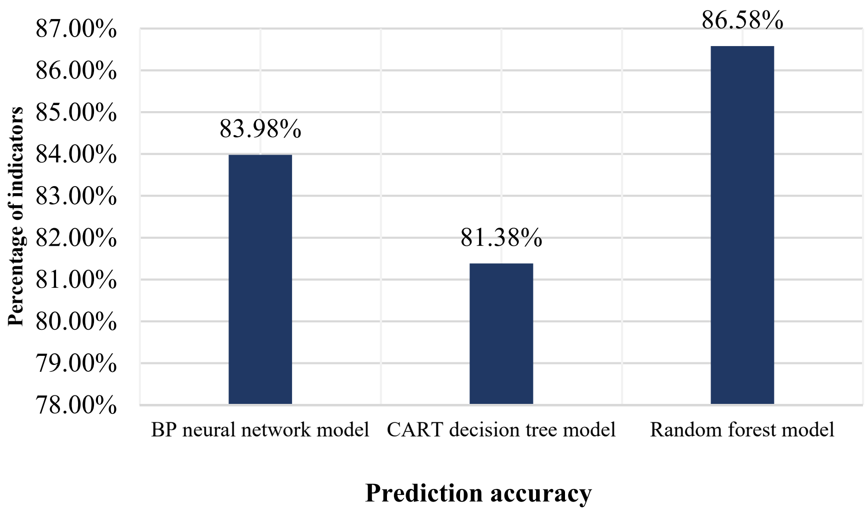

Figure 7. Comparison of model prediction accuracy.

Detailed contrasts are conducted based on data from Figure 7. The recognition and prediction performance of the random forest model on the performance level of the fuel consumption of HDDTs is enhanced and improved compared with the BP neural network model and CART decision tree model. The prediction accuracy of the CART decision tree model is only 81.38%, which is far lower than that of the random forest model. This is well understood. Compared with the decision tree model, the latter can better avoid overfitting, resulting in improved prediction accuracy. In addition, the partial lack of data is also an important reason for the low prediction accuracy of the model. The results fully express that the random forest model is more suitable as a mathematical model established in this research to identify, classify and predict the fuel consumption of HDDTs.

Based on the above analysis results, this paper chooses a more accurate method for predicting the fuel consumption of HDDTs, namely the random forest model. The selection of this method has practical significance, that is, transportation enterprises can accurately predict the daily fuel consumption of HDDTs in the future by using the established random forest model. Combined with the analysis results of the binary Logistic regression model, they can focus on training drivers' driving behavior to reduce the fuel consumption of HDDTs.

4. Conclusions

This paper showed how to make full use of the driving data collected by the fleet management system to analyze the fuel consumption of HDDTs. This paper used fuel consumption per 100 km of HDDTs as the dependent variable and extracted 21 influencing factors from the human-vehicle-road-environment system as independent variables. These independent variables involved four aspects: vehicle-related, environment-related, driving-related, and road-related factors, which were combined to explore the generation mechanism of different fuel consumption during driving.

This study combined traditional statistical methods with machine learning methods. An explanatory traditional binary Logistic regression model was used to analyze the fuel consumption data of HDDTs. The model quantified the influential degree of each significant factor on fuel consumption and the probability of fuel consumption performance. For example, if the engine revolution speed is too high, the fuel consumption of HDDTs may also increase, which requires the drivers to try to avoid driving at low gear and high speed. When the outdoor temperature is high, the fuel consumption will increase accordingly, etc. According to the machine learning method, the BP neural network model, CART decision tree model, and random forest model emphasized predictability were established to realize the classification prediction of fuel consumption. These models were compared and verified by outputting the prediction accuracy, and the research result highlighted the superiority of the established random forest model for the prediction of fuel consumption. The prediction accuracy of the random forest model could reach 86.58%, which could accurately predict the occurrence of high fuel consumption classification during actual driving. The main conclusions of this paper can be summarized as follows: first, the HDDTs load, average engine speed, average speed, altitude, temperature and precipitation during driving are significantly related to the fuel consumption of HDDTs; second, a random forest model may be the most appropriate method to predict fuel consumption of HDDTs in the real driving environment; third, transportation enterprises with fleet management system can control the driving fuel consumption more strictly according to the arguments put forward in this paper.

Based on the realistic environment and actual data, this paper explored the generating mechanism of different fuel consumption performances and the effect of multi-dimensional factors on fuel consumption. It could provide analytical method reference and data support for predicting, evaluating, and managing the fuel consumption of HDDTs in medium and long-distance transportation. This research could also help to realize the goal of reducing cost and increasing efficiency of vehicle driving, and made particular contributions to control traffic pollution and reduce exhaust gas.

Due to the cutting-edge and methodological complexity of data mining, there is much room for further development in this study. In future research, under the premise of ensuring the reasonableness and accuracy of data, the data collected will be expanded in type and quantity. The model prediction accuracy will be improved by trying other data mining methods. This research only focused on the mining and analysis of the fuel consumption data. In the future, according to the expansion of the sample size, based on the prediction model and the extracted significant factors, different indicators can be weighted. The fuel consumption evaluation system of HDDTs can be established to realize the scientific management of the fuel consumption of HDDTs.

Author Contributions: Conceptualization, L.L.; Data curation, J.G.; Methodology, L.L.; Software, J.G. and J.S.; Writing-original draft preparation, J.G. and J.S.; Review and editing, C.Z. and J.M.; Supervision, J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (Grant No. 51778141 and 52072069), Jiangsu Creative PhD student sponsored project (KYCX20_00138) and Transportation Department of Henan Province (sponsored by project 2018G7). And the APC was funded by National Natural Science Foundation of China (Grant No. 52072069).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from TopChains International Logistics Co., Ltd. and are available from the author Jian Gong with the permission of TopChains International Logistics Co., Ltd.

Acknowledgments: The authors of this paper would like to thank TopChains International Logistics Co., Ltd. for providing the statistical data used in this research. This research was funded by National Natural Science Foundation of China (Grant No. 51778141 and 52072069), Jiangsu Creative PhD student sponsored project (KYCX20_00138) and Transportation Department of Henan Province (sponsored by project 2018G7). Their assistance is gratefully acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Guensler, R.; Liu, H.; Xu, Y.; Akanser, A.; Kim, D.; Hunter, M.P.; Rodgers, M.O. Energy consumption and emissions modeling of individual vehicles. *Transp. Res. Rec.* **2017**, *2627*, 93–102. [CrossRef]
2. Ahn, K.; Rakha, H.; Trani, A.; Van Aerde, M. Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels. *J. Transp. Eng.* **2002**, *128*, 182–190. [CrossRef]
3. Hlasny, T.; Fanti, M.P.; Mangini, A.M.; Rotunno, G.; Turchiano, B. Optimal fuel consumption for heavy trucks: A review. In Proceedings of the 2017 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI 2017), Bari, Italy, 18–20 September 2017; pp. 80–85. [CrossRef]
4. Li, D.; Li, C.; Miwa, T.; Morikawa, T. An exploration of factors affecting drivers' daily fuel consumption efficiencies considering multi-level random effects. *Sustainability* **2019**, *11*, 393. [CrossRef]
5. Chen, Y.; Zhu, L.; Gonder, J.; Young, S.; Walkowicz, K. Data-driven fuel consumption estimation: A multivariate adaptive regression spline approach. *Transp. Res. Part C Emerg. Technol.* **2017**, *83*, 134–145. [CrossRef]
6. Ericsson, E. Independent driving pattern factors and their influence on fuel-use and exhaust emission factors. *Transp. Res. Part D Transp. Environ.* **2001**, *6*, 325–345. [CrossRef]
7. Zhou, M.; Jin, H.; Wang, W. A review of vehicle fuel consumption models to evaluate eco-driving and eco-routing. *Transp. Res. Part D Transp. Environ.* **2016**, *49*, 203–218. [CrossRef]
8. Cachón, L.; Pucher, E. Fuel Consumption Simulation Model of a CNG Vehicle based on Real-world Emission Measurement. *SAE Tech. Pap.* **2007**, *24*, 114–123. [CrossRef]
9. Wang, H.; Fu, L.; Zhou, Y.; Li, H. Modelling of the fuel consumption for passenger cars regarding driving characteristics. *Transp. Res. Part D Transp. Environ.* **2008**, *13*, 479–482. [CrossRef]
10. Xiang, Y.; Li, Z.; Chen, M.Y.; Jiang, X.M. Research on calculation software of fuel consumption for heavy trucks. In Proceedings of the 3rd International Conference on Measuring Technology and Mechatronics Automation (ICMTMA 2011), Shanghai, China, 6–7 January 2011; Volume 2, pp. 1121–1124. [CrossRef]
11. Wang, J.; Rakha, H.A. Fuel consumption model for heavy duty diesel trucks: Model development and testing. *Transp. Res. Part D Transp. Environ.* **2017**, *55*, 127–141. [CrossRef]
12. Ma, H.; Xie, H.; Chen, S.; Yan, Y.; Huang, D. Effects of driver acceleration behavior on fuel consumption of city buses. *SAE Tech. Pap.* **2014**, *1*, 389–395. [CrossRef]
13. Du, Y.; Wu, J.; Yang, S.; Zhou, L. Predicting vehicle fuel consumption patterns using floating vehicle data. *J. Environ. Sci.* **2017**, *59*, 24–29. [CrossRef] [PubMed]
14. Wysocki, O.; Deka, L.; Elizondo, D. Heavy Duty Vehicle Fuel Consumption Modelling Using Artificial Neural Networks. In Proceedings of the 25th International Conference on Automation and Computing (ICAC), Lancaster, UK, 5–7 September 2019; pp. 1–6. [CrossRef]
15. Mane, A.; Djordjevic, B.; Ghosh, B. A data-driven framework for incentivising fuel-efficient driving behaviour in heavy-duty vehicles. *Transp. Res. Part D Transp. Environ.* **2021**, *95*, 102845. [CrossRef]
16. Walnum, H.J.; Simonsen, M. Does driving behavior matter? An analysis of fuel consumption data from heavy-duty trucks. *Transp. Res. Part D Transp. Environ.* **2015**, *36*, 107–120. [CrossRef]
17. Toledo, G.; Shiftan, Y. Can feedback from in-vehicle data recorders improve driver behavior and reduce fuel consumption? *Transp. Res. Part A Policy Pract.* **2016**, *94*, 194–204. [CrossRef]
18. Sun, R.; Chen, Y.; Dubey, A.; Pugliese, P. Hybrid electric buses fuel consumption prediction based on real-world driving data. *Transp. Res. Part D Transp. Environ.* **2021**, *91*, 102637. [CrossRef]
19. Guo, D.; Wang, J.; Zhao, J.B.; Sun, F.; Gao, S.; Li, C.D.; Li, M.H.; Li, C.C. A vehicle path planning method based on a dynamic traffic network that considers fuel consumption and emissions. *Sci. Total Environ.* **2019**, *663*, 935–943. [CrossRef] [PubMed]
20. Faria, M.V.; Duarte, G.O.; Varella, R.A.; Farias, T.L.; Baptista, P.C. How do road grade, road type and driving aggressiveness impact vehicle fuel consumption? Assessing potential fuel savings in Lisbon, Portugal. *Transp. Res. Part D Transp. Environ.* **2019**, *72*, 148–161. [CrossRef]

21. Lee, S. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *Int. J. Remote Sens.* **2005**, *26*, 1477–1491. [CrossRef]
22. Ozdemir, A. Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey). *J. Hydrol.* **2011**, *405*, 123–136. [CrossRef]
23. Zhang, Z.; Wang, C. The Research of Vehicle Plate Recognition Technical Based on BP Neural Network. *AASRI Procedia* **2012**, *1*, 74–81. [CrossRef]
24. Ma, J.; Li, D.; Lu, Y.; Chen, J. Intelligent Diagnosis System for Vehicle Network Based on BP Neural Network. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, *452*, 42004. [CrossRef]
25. Yao, Y.; Zhao, X.; Liu, C.; Rong, J.; Zhang, Y.; Dong, Z.; Su, Y.; Chen, F. Vehicle Fuel Consumption Prediction Method Based on Driving Behavior Data Collected from Smartphones. *J. Adv. Transp.* **2020**, *2020*, 9263605. [CrossRef]
26. Li, H.; Dong, H.; Jia, L.; Ren, M. Vehicle classification with single multi-functional magnetic sensor and optimal MNS-based CART. *Meas. J. Int. Meas. Confed.* **2014**, *55*, 142–152. [CrossRef]
27. Zeng, L.; Guo, J.; Wang, B.; Lv, J.; Wang, Q. Analyzing sustainability of Chinese coal cities using a decision tree modeling approach. *Resour. Policy* **2019**, *64*, 101501. [CrossRef]
28. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
29. Zhang, X.; Zhang, J.; Zhang, J.; Zhang, Y. Research on the Combined Prediction Model of Residential Building Energy Consumption Based on Random Forest and BP Neural Network. *Geofluids* **2021**, *2021*, 7271383. [CrossRef]
30. Li, L.H.; Zhu, J.S.; Shan, X.H.; Zhang, X. Prediction modeling of railway short-term passenger flow based on random forest regression. In *Lecture Notes in Electrical Engineering*; Springer: Singapore, 2019; Volume 503. [CrossRef]
31. Wang, Y.; Zhang, J.; Liu, P.; Xu, C.; Zhang, F. Identification of Non-Green Channel Vehicles at Highway Toll Gate Based on Logistic Regression Model. In Proceedings of the 19th COTA International Conference of Transportation Professionals, Nanjing, China, 6–8 July 2019. [CrossRef]
32. Schendera, C.F. *Regressionsanalyse Mit SPSS*; De Gruyter Oldenbourg: Berlin, Germany, 2014.
33. Peng, Y.; Peng, S.; Wang, X.; Tan, S. An investigation on fatality of drivers in vehicle–fixed object accidents on expressways in China: Using multinomial logistic regression model. *Proc. Inst. Mech. Eng. Part H J. Eng. Med.* **2018**, *232*, 643–654. [CrossRef] [PubMed]
34. Zargarnezhad, S.; Dashti, R.; Ahmadi, R. Predicting vehicle fuel consumption in energy distribution companies using ANNs. *Transp. Res. Part D Transp. Environ.* **2019**, *74*, 174–188. [CrossRef]
35. Dobre, A. Theoretical Study Regarding the Fuel Consumption Performance for a Vehicle Equipped with a Mechanical Transmission. *Procedia Manuf.* **2019**, *32*, 537–544. [CrossRef]
36. Zhou, X.; Huang, J.; Lv, W.; Li, D. Fuel consumption estimates based on driving pattern recognition. In Proceedings of the 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, Beijing, China, 20–23 August 2013; pp. 496–503. [CrossRef]
37. Hao, C.; Ge, Y.; Wang, X. Heavy-duty diesel engine fuel consumption comparison with diesel and biodiesel measured at different altitudes. *Int. J. Veh. Perform.* **2020**, *6*, 263–276. [CrossRef]
38. Jeon, H. The impact of climate change on passenger vehicle fuel consumption: Evidence from U.S. panel data. *Energies* **2019**, *12*, 4460. [CrossRef]
39. Soni, A.R.; Chandel, M.K. Impact of rainfall on travel time and fuel usage for Greater Mumbai city. *Transp. Res. Procedia* **2020**, *48*, 2096–2107. [CrossRef]
40. He, J.; Qi, Z.; Hang, W.; Zhao, C.; King, M. Predicting freeway pavement construction cost using a back-propagation neural network: A case study in Henan, China. *Baltic J. Road Bridge Eng.* **2014**, *9*, 66–76. [CrossRef]
41. Lantz, B. *Machine Learning with R*, 2nd ed.; Packt Publishing: Birmingham, UK, 2015.
42. Trigila, A.; Iadanza, C.; Esposito, C.; Scarascia-Mugnozza, G. Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). *Geomorphology* **2015**, *249*, 119–136. [CrossRef]
43. Zhou, X.; Lu, P.; Zheng, Z.; Tolliver, D.; Keramati, A. Accident Prediction Accuracy Assessment for Highway-Rail Grade Crossings Using Random Forest Algorithm Compared with Decision Tree. *Reliab. Eng. Syst. Saf.* **2020**, *200*, 106931. [CrossRef]